iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** V **Month of publication:** May 2026

DOI: https://doi.org/10.22214/ijraset.2026.82739

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

Detection of Credit Card Frauds Using Machine Learning Techniques

Riya Wagh[1], Reena Kharat[2]
Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, India

Abstract: ***The popularity of credit cards in the modern world is associated with the growing rate of credit card fraud. Credit card technology has made e-shopping easier, but at the same time increased the possibility of fraudulent activities. Computer algorithms such as machine learning can help in identifying fraudulent activities. Machine learning is very effective in analyzing customer information. Recently, the number of credit card fraud cases has been constantly growing and has caused significant losses for customers, retailers, and banks. In the proposed research paper, various methods of identification of credit card fraud detection via machine learning will be analyzed and compared according to their effectiveness.***
Keywords: Credit card frauds, Machine Learning, Random Forest Algorithm, Artificial Neural Network (ANN).

I. INTRODUCTION

There are several reasons why credit cards are preferred methods for purchasing goods online. Firstly, it is very convenient for the buyers. In addition, the number of credit card fraud cases and abuse increases due to an increase in the number of purchases using these cards. There are some problems related to credit card fraud which arise in the course of such activity. The perpetrators of such crimes are becoming more and more resourceful and can find ways of stealing the money without any chance to be detected. There are different types of credit card fraud such as the unauthorized usage of a card, performing strange transactions, and operating an expired credit card. The increase in credit card fraud cases occurs more and more often in recent years[1]The aim of this study is to provide effective ways to fight against the cases mentioned above. We will consider such aspects as publicly available information, imbalance of data (one particular data type is significantly bigger than the other), the dynamic nature of the problem, and many false alerts. The key element distinguishing OBeP from any other internet payments is the presence of the real-time customer verification by means of the online banking facility. Due to the use of the technology mentioned and the rapid development of IT facilities around the globe, credit cards have become the most crucial means of payment nowadays. The sustainability of the entire e-payment system relies significantly on the enhancement of measures for detection of fraud. In case with credit card fraud detection, there exist two categories of transactions – those which are fraudulent and those which are legitimate [3]. One of the most effective measures aimed at detecting fraud is machine learning technology. Machine learning techniques can be divided into two groups. The first type includes supervised learning, when the model has been taught by examples labeled as fraudulent. The second one implies grouping the clients according to their spending habits using unsupervised learning.

II. LITERATURE REVIEW

Several methods were employed by Zareapoor and his team to identify the most effective model for identifying fraud. This is based on the model accuracy, detection speed, and costs involved [25]. The model adopted by Alenzi and Aljehane in detecting fraud in credit cards was the Logistic Regression. The score obtained from their model was 97.2% accuracy, 97% sensitivity, and 2.8% error rate. The performance of the model was compared with those of two others; the 5 Voting Classifiers and KNN [2].

The model developed by Maniraj was able to classify whether the new transaction is a fraud or not fraud, their objective was to achieve 100% in the detection of fraudulent transactions apart from minimizing the wrongly classified fraudulent instances. This has been successful for their model as they have managed to reach 99.7% fraudulent transaction detections [17]. The classification technique employed by Dheepa and Dhanapal was behavior-based classification technique which involved the use of SVM to classify credit card fraud based on the customer's behavior including the amount, date, time, location, and frequency of card utilization. This approach led to an accuracy of over 80% [7]. Mailini and Pushpa proposed the use of KNN and outlier detection techniques in the detection of credit card fraud. The authors concluded through their simulation tests that KNN is the appropriate approach for detection and determination of anomaly in a target instance with memory limitations [16].

©IJRASET: All Rights are Reserved

Maes et al., on the other hand, suggested Bayesian and Neural Network as a method in the identification of credit card frauds. Their findings indicated that the Bayesian algorithm is 8% more efficient compared to ANN in fraud detection; this suggests that at times BBN detects 8% more fraud transactions[14]. As regards the accuracy in 10:90 distribution, it appears that the most accurate algorithms are Naïve Bayes with 97.5%, followed by KNN with 97.1%, and logistic regression algorithm with 36.4%. On another note, another interesting distribution is the 34:66 distribution; in this case, the highest accuracy obtained is by the KNN with 97.9%, followed by Naïve Bayes with 97.6%, while the accuracy of logistic regression increased to 54.8% [3]. Jain et al. have employed several machine learning algorithms to differentiate credit card fraud, such as SVM, ANN, and KNN. To make a comparison between the models, they have computed TP, FN, FP, and TN rates. ANN accuracy is 99.71%, precision is 99.68%, and false alarm rate is 0.12%. SVM's accuracy is 94.65%, precision is 85.45%, and the false alarm rate is 5.2%.

The research conducted by Gupta et al. aimed at implementing an automated model that applies a combination of ML approaches in order to identify the economic links between the user and the fraudulent instances. This approach has been narrowed down to credit card transaction analysis. The most successful method of distinguishing fraudulent transactions out of all approaches is the Naïve Bayes algorithm, which demonstrated accuracy equal to 80.4% and 96.3% area under the curve [10]. Adepoju et al. utilized all ML techniques described in the paper, including Logistic Regression, SVM, Naïve Bayes, and KNN algorithms. The distorted credit card fraud data was analyzed using these techniques, and their performance was assessed based on accuracy scores: 99.07% for Logistic Regression, 95.98% for Naïve Bayes, 96.91% for K-nearest neighbor, and 97.53% for SVM [1].Safa and Ganga explored the effectiveness of Logistic Regression, (KNN) K-Nearest Neighbor, and Naive Bayes on extremely corrupted credit card dataset. They performed their experiments in Python language; the best method was decided based on evaluations. The accuracy rate of their proposed model for Naive Bayes is 83%, for Logistic Regression is 97.69%, while for K-Nearest Neighbor it is 54.86% [19]. In the case study by the group of Varmedja, various machine learning models were used including Logistic Regression, Multilayer Perceptron, Random Forest, and Naive Bayes. Since the data used was highly imbalanced, the authors applied the SMOTE algorithm to increase the sample size, feature selection, as well as splitting the data into training and testing data sets. The highest scoring model among all experiments was Random Forest with 99.96%, and slightly lower is Multilayer Perceptron with 99.93%. Naive Bayes is next with 99.23%, and finally Logistic Regression with 97.46% [24].

The algorithm for identifying credit card fraud was developed by Sailusha and his associates and its primary purpose is to identify any fraud within the process. Some of the algorithms used by their model include adaboost and Random Forest; accuracy of adaboost is 99.90%. This means that adaboost outperformed Random Forest in terms of accuracy, which stands at 93.99%. Naïve Bayes had better performance than K-nearest neighbor as it was able to achieve an accuracy of 95%, while KNN scored only 90%[13]. Najdat and his colleagues' algorithm for the detection of credit card fraud transactions includes bi-directional Long short-term memory (BiLSTM) as well as (BiGRU) bi-directional Gated recurrent unit. Additionally, the group chose to implement six machine learning classifiers, which include Voting, Adaboost, Random Forest, Decision Tree, Naïve Bayes, and Logistic Regression. Accuracy scores of these machine learning algorithms stand at 99.13% and 96.27% for K-nearest neighbor and logistic regression respectively, 96.40% for decision tree and 96.98% for Naïve Bayes[18].

Saheed et al.'s paper deals with detection of Credit Card Fraud using the (GA) Genetic Algorithm for selecting features. Here, in feature selection, the dataset is divided into two categories, namely, first priority features and second priority features. For implementing the above algorithm, Saheed et al. selected three machine learning techniques, namely, The Naïve Bayes (NB), Random Forest (RF), and Support Vector Machine (SVM). The accuracies achieved through the Naïve Bayes, Support Vector Machine, and Random Forest algorithms were 94.3%, 96.3%, and 96.40%, respectively [20]. The research conducted by Itoo et al. involved three machine learning techniques, namely, Logistic Regression, Naïve Bayes, and K-Nearest Neighbors. These researchers have documented the research process and comparative analysis performed. Their experiments were coded using Python. Here, the accuracy values of Logistic Regression, Naïve Bayes, and K-Nearest Neighbors were 91.2%, 85.4%, and 66.9%, respectively [11].

Tanouz et al. considered exploring different classification algorithms such as Naïve Bayes, Logistic Regression, Random Forest, and Decision Trees for dealing with extremely imbalanced datasets. In addition, they also included the calculation of five different performance measures including accuracy, precision, recall, confusion matrix, and ROC-AUC score[23]. Dighe and his collaborators made use of four different machine learning classifiers such as KNN, Naïve Bayes, Logistic Regression, and Neural Networks, Multi-layered Perceptron, and Decision Trees for their study and evaluated their performance using several accuracy measures. The Naïve Bayes classifier achieved an accuracy of 96.98%, the third best-performing classifier 96.40%, and the last one was the logistic regression classifier at 96.27%[8].

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Conclusion about the classifier of the project based on the study done in their paper. The logistic regression model showed an accuracy rate of 99.8%, while random forest obtained an accuracy of 100% and naïve bayes had an accuracy rate of 90.8%. According to Sahin and Duman, there are four ways of using a support vector machine in detection of credit card fraud. These include Support Vector Machine with RBF, Polynomial, Sigmoid, and Linear Kernel models which all had an accuracy rate of 99.87% for the training set and 83.02% for testing [21].

III. PROBLEM STATEMENT

A. *Problem Statement*

The use of machine learning techniques to deal with credit card fraud is inherently hindered due to the presence of extremely imbalanced classes of target variables, which results in false-positive detections and late detection times. This research tries to address this issue by employing a Random Forest algorithm that has been optimized through an algorithm pipeline for customization and tuning purposes.

B. *Objectives*

a. Reduce Financial Losses to Both the Credit Card Company and Its Customers:
Develop a real-time system for detecting frauds, enabling them to minimize financial losses for both the credit card company and its customers.
b. Develop a User-Friendly System:
Make the system user-friendly by developing an easily navigable and simple interface.
c. Design a System That Is Easily Maintainable:
Develop a system in such a way that, in case of any glitches in the system, they can be easily located and corrected. This can be done by designing the system in a modular manner.
d. Detect Frauds More Quickly:
Develop a real-time system for detecting frauds.

C. *Algorithms for Credit Card Fraud Detection*

1) Random Forest: Random Forest is a well-known algorithm that falls under the domain of machine learning and can be applied to both classification and regression problems. The Random Forest algorithm comes under the family of ensemble learning techniques.
2) Support Vector Machine: Support Vector Machine (SVM) is a supervised machine learning model that performs well in both classification and regression problems. The SVM model finds the optimal hyperplane to separate different classes. From the Confusion Matrix, we can compute the following measures:

a) Precision:
Precision refers to the precision of positive classifications made by the machine learning model during classification problems like detecting fraud in credit card transactions.
Mathematically, precision is defined as:

$$Precision = \frac{True\ positive}{True\ positive + False\ positive}$$

b. Recall:
Also referred to as sensitivity or true positive rate, recall is an evaluation measure used to indicate how complete the predictions generated by the model are.
Mathematically, recall is defined as:

$$Recall = \frac{True\ positive}{True\ positive + false\ positive}$$

c. F1-score:
The F1 score is a performance metric that integrates both precision and recall into one single number.
Mathematically, the F1-score is calculated as:

$$F1\ score = \frac{2xPrecisionxRecallxPrecision}{Recall}$$

d. Accuracy:

Accuracy is an evaluation measure indicating how accurate the model's predictions are.

Mathematically, recall is defined as:

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions}$$

IV. PROPOSED METHODOLOGY

Machine learning algorithms are used by credit card fraud detection systems to examine vast amounts of transaction data and spot possibly fraudulent activity. This is how they typically operate:

A. Data Collection

The intake phase maintains a persistent stream from the log files of the credit card network to collect all incoming transaction data vectors. Data vectors comprise essential attributes such as quantitative measures (transaction cost, timestamps), contextual environmental factors (merchant type codes, regional billing terminals) and cardholder attributes (geographical regions, expenditure behavior, account characteristics).

B. Data Preprocessing

The raw input data is subjected to data cleaning and transformations before being modeled on computers. Data cleaning in this process involves replacing missing data using statistical methods such as mean imputation, removing duplicate entries, mapping non-numerical attributes using categorical mapping, and normalizing high values through Z-score or Min-Max normalization to avoid the influence of large metric values on distance calculations.

C. Feature Engineering

Feature Engineering uses feature discovery and correlation of behaviors to find patterns. In case of imbalanced data where real actions exceed fraudulent actions, SMOTE is applied on the training dataset in order to develop synthetic borderline regions in respect of the minority class, instead of over-fitted replicas.

D. Model Training

The balanced data sets are divided into 70% for training and 30% for verification pools to develop a reliable multiple model classification framework. The supervised classifiers used are random forest (decision tree ensembles), support vector machines (hyperplanes in high dimensions), logistic regression (sigmoid function fitting), and artificial neural networks (neural network nodes using backpropagation).

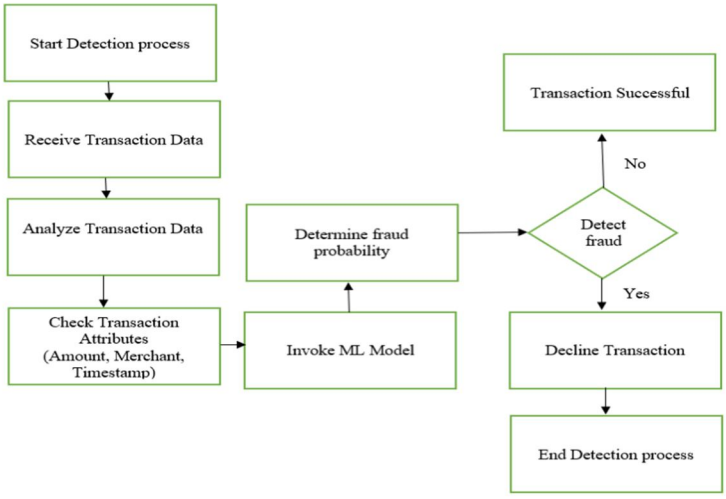


Fig 4.1: System flow

Above fig 5.1 System workflow runs as a real-time risk management pipeline, which handles credit card transactions and attempts to prevent fraud prior to any transaction authorization. Once a transaction trigger occurs, the system absorbs raw transaction data and analyses it to identify key attributes such as the transaction amount, merchant information, and timestamps. Based on the variables, the machine learning model predicts the likelihood of fraud occurring, based on past spending behaviors. In case there is no anomaly detected by the decision gate, the transaction succeeds; otherwise, the transaction fails to mitigate financial risks.

E. Evaluation Matrix

The metrics of the system are evaluated by using a confusion matrix, which includes the values of True Positive (blocked cases of fraud), True Negative (approved cases of actual spending), False Positive (erroneous positives), and False Negatives (uncaught leaks). The above metrics lead to scoring metrics that include Precision, which evaluates positive predictions accuracy; Recall, which evaluates system completeness; F1-Score, which balances both metrics; and Accuracy.

V. RESULT

Analysis of the supervised learning algorithms shows that the Random Forest Classifier performed better than all other algorithms having a validation accuracy score of 95.94%, thus being a powerful tool for tackling any complex patterns within the dataset and handling the problem of class imbalance. Meanwhile, the SVM is able to separate classes in space effectively but suffers from higher errors in the form of 5.2% false positives. As for the Logistic Regression and Naïve Bayes Classifier, which are considered parametric baselines, they performed accurately yet worse than the ensemble classifier yielding scores of 91.2% and 85.4% respectively. However, it must be noted that despite their efficiency, these models are not able to describe the non-linear behavior associated with the current credit card frauds.

Table 5.1 Result Comparison

Machine Learning Model Variant	Observed Classification Accuracy
Random Forest Classifier	95.94%
Support Vector Machine (Linear SVM)	94.65%
Logistic Regression	91.2%
Naïve Bayes Classifier	85.4%

The table 5.1 shows four different types of supervised learning methods evaluated on the basis of their accuracies of classification of credit card frauds. The method that provides the maximum accuracy of 95.94% is that of the Random Forest Classifier, using the property of its ensemble tree method to deal with the complexities and huge imbalance in the data. Next comes Linear SVM having an accuracy of 94.65%; it gives good class separation but has relatively high levels of false alarms. Parametric approaches are comparatively poorer performers, with Logistic Regression providing 91.2% and the Naïve Bayes Classifier 85.4%.

VI. CONCLUSION

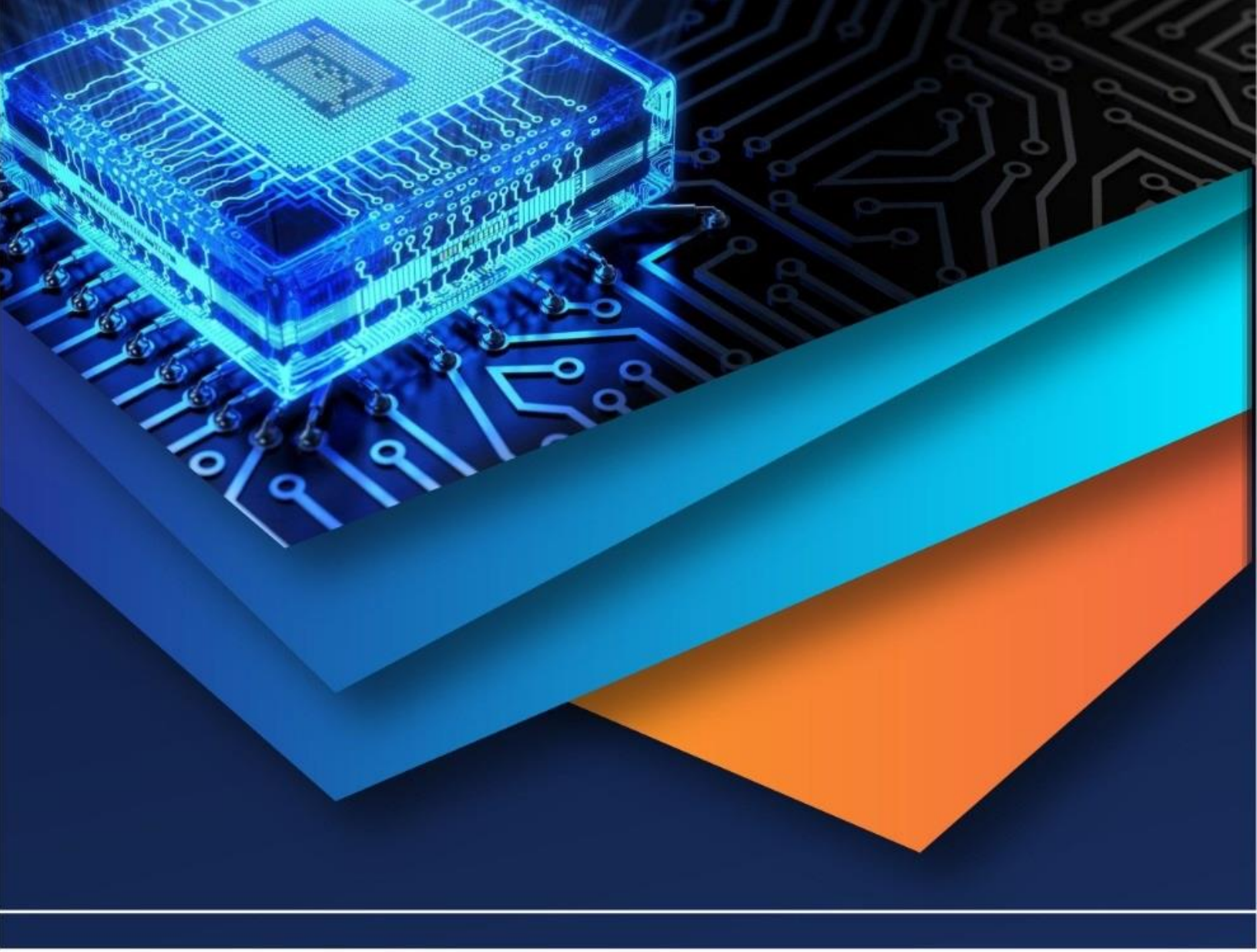
In conclusion, the main objective of this project is to find the most suited model for credit card fraud detection in terms of the machine learning techniques chosen for the project, Explored various computer methods for detecting fraudulent credit card transactions. Evaluated performance using metrics like accuracy, precision, and recall. Choose Random Forest Algorithm as the preferred method which scored 95.94% accuracy. Described it as a smart helper for identifying suspicious transactions. It emphasized the goal of ensuring the security of financial transactions. I believe that using this model will help to decrease the amount of credit card fraud and increase the customer's satisfaction as it will provide them with a better experience in addition to feeling secure.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

REFERENCES

[1] Adepoju, O., Wosowei, J., lawte, S., & Jaiman, H. (2019). Comparative evaluation of credit card fraud detection using machine learning techniques. 2019 Global Conference for Advancement in Technology (GCAT). https://doi.org/10.1109/gcat47503.2019.8978372
[2] Alenzi, H. Z., & Aljehane, N. O. (2020). Fraud detection in credit cards using logistic regression. International Journal of Advanced Computer Science and Applications, 11(12). https://doi.org/10.14569/ijacsa.2020.0111265
[3] Awoyemi, J. O., Adetunmbi, A. O., & Oluwadare, S. A. (2017). Credit card fraud detection using Machine Learning Techniques: A Comparative Analysis. 2017 International Conference on Computing Networking and Informatics (ICCNI). https://doi.org/10.1109/iccni.2017.8123782
[4] Bhanusri, A., Valli, K. R. S., Jyothi, P., Sai, G. V., & Rohith, R. (2020). Credit card fraud detection using Machine learning algorithms. Journal of Research in Humanities and Social Science, 8(2), 04-11
[5] Credit card statistics. Shift Credit Card Processing. (2021, August 30). Retrieved from https://shiftprocessing.com/credit-card/
[6] Daly, L. (2021, October 27). Identity theft and credit card fraud statistics for 2021: The ascent. The Motley Fool. Retrieved from https://www.fool.com/the-ascent/research/identity-theft-credit-card-fraud-statistics/
[7] Dheepa, V., & Dhanapal, R. (2012). Behavior based credit card fraud detection using support vector machines. ICTACT Journal on Soft Computing, 02(04), 391–397. https://doi.org/10.21917/ijsc.2012.0061 24
[8] Dighe, D., Patil, S., & Kokate, S. (2018). Detection of credit card fraud transactions using machine learning algorithms and Neural Networks: A comparative study. 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA). https://doi.org/10.1109/iccubea.2018.8697799
[9] Domínguez-Almendros, S., Benítez-Parejo, N., & Gonzalez-Ramirez, A. R. (2011). Logistic regression models. Allergologia et immunopathologia, 39(5), 295-305.
[10] Gupta, A., Lohani, M. C., & Manchanda, M. (2021). Financial fraud detection using naive Bayes algorithm in highly imbalance data set. Journal of Discrete Mathematical Sciences and Cryptography, 24(5), 1559–1572. https://doi.org/10.1080/09720529.2021.1969733
[11] Itoo, F., Meenakshi, & Singh, S. (2020). Comparison and analysis of logistic regression, Naïve Bayes and Knn Machine Learning Algorithms for credit card fraud detection. International Journal of Information Technology, 13(4), 1503–1511. https://doi.org/10.1007/s41870-020-00430-y
[12] Jain, Y., NamrataTiwari, S., & Jain, S. (2019). A comparative analysis of various credit card fraud detection techniques. International Journal of Recent Technology and Engineering, 7(5S2), 402-407
[13] Kiran, S., Guru, J., Kumar, R., Kumar, N., Katariya, D., & Sharma, M. (2018). Credit card fraud detection using Naïve Bayes model based and KNN classifier. International Journal Of Advance Research, Ideas And Innovations In Technology, 4(3).
[14] Maes, S., Tuyls, K., Vanschoenwinkel, B., & Manderick, B. (2002, January). Credit card fraud detection using Bayesian and neural networks. In Proceedings of the 1st nternational naiso congress on neuro fuzzy technologies (pp. 261-270).
[15] Mahesh, B. (2020). Machine Learning Algorithms - A Review, 9(1). https://doi.org/10.21275/ART20203995 25
[16] Malini, N., & Pushpa, M. (2017). Analysis on credit card fraud identification techniques based on KNN and outlier detection. 2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB). https://doi.org/10.1109/aeeicb.2017.7972424
[17] Maniraj, S. P., Saini, A., Ahmed, S., & Sarkar, S. D. (2019). Credit card fraud detection using machine learning and Data Science. Credit Card Fraud Detection Using Machine Learning and Data Science, 08(09). https://doi.org/10.17577/ijertv8is090031
[18] Najadat, H., Altiti, O., Aqouleh, A. A., & Younes, M. (2020). Credit card fraud detection based on machine and Deep Learning. 2020 11th International Conference on Information and Communication Systems (ICICS). https://doi.org/10.1109/icics49469.2020.239524
[19] Safa, M. U., & Ganga, R. M. (2019). Credit Card Fraud Detection Using Machine Learning. International Journal of Research in Engineering, Science and Management, 2(11).
[20] Saheed, Y. K., Hambali, M. A., Arowolo, M. O., & Olasupo, Y. A. (2020). Application of ga feature selection on Naive Bayes, random forest and SVM for credit card fraud detection. 2020 International Conference on Decision Aid Sciences and Application (DASA). https://doi.org/10.1109/dasa51403.2020.9317228
[21] Sahin, Y., & Duman, E. (2011). Detecting Credit Card Fraud by Decision Trees and Support Vector Machines. Proceedings of the International MultiConference of Engineers and Computer Scientists, 1.
[22] Sailusha, R., Gnaneswar, V., Ramesh, R., & Rao, R. R. (n.d.). Credit Card Fraud Detection Using Machine Learning. Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS 2020). 26
[23] Tanouz, D., Subramanian, R. R., Eswar, D., Reddy, G. V., Kumar, A. R., & Praneeth, C. H. V. (2021). Credit card fraud detection using machine learning. 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS). https://doi.org/10.1109/iciccs51141.2021.9432308
[24] Varmedja, D., Karanovic, M., Sladojevic, S., Arsenovic, M., & Anderla, A. (2019). Credit Card Fraud Detection - machine learning methods. 2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH). https://doi.org/10.1109/infoteh.2019.8717766
[25] Zareapoor, M., Seeja.K.R, S. K. R., & Afshar Alam, M. (2012). Analysis on credit card fraud detection techniques: Based on certain design criteria. International Journal of Computer Applications, 52(3), 35–42. https://doi.org/10.5

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)